MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 6: 75–87
DOI 10.3897/mbmg.6.78756

Emerging Technique

A two-step metagenomics approach for the identification and mitochondrial DNA contig assembly of vertebrate prey from the blood meals of common vampire bats (*Desmodus rotundus*)

Physilia Y. S. Chua[1,2], **Christian Carøe**[1], **Alex Crampton-Platt**[3], **Claudia S. Reyes-Avila**[1], **Gareth Jones**[4], **Daniel G. Streicker**[5,6], **Kristine Bohmann**[1]

1 *Section for Evolutionary Genomics, GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, 1353 Copenhagen K, Denmark*

2 *Department of Biology, Faculty of Science, University of Copenhagen, 1353 Copenhagen K, Denmark*

3 *NatureMetrics, 1 Occam Court, Surrey Research Park, Guildford, GU2 7HJ, UK*

4 *School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK*

5 *Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK*

6 *MRC – University of Glasgow Centre for Virus Research, Glasgow G61, 1QH, UK*

Corresponding author: Physilia Y. S. Chua (physiliachua@gmail.com)

Academic editor: Carmelo Andújar | **Received** 6 December 2021 | **Accepted** 14 March 2022 | **Published** 7 April 2022

Abstract

The feeding behaviour of the sanguivorous common vampire bat (*Desmodus rotundus*) facilitates the transmission of pathogens that can impact both human and animal health. To formulate effective strategies in controlling the spread of diseases, there is a need to obtain information on which animals they feed on. One DNA-based approach, shotgun sequencing, can be used to obtain such information. Even though it is costly, shotgun sequencing can be used to simultaneously retrieve prey and vampire bat mitochondrial DNA for population studies within one round of sequencing. However, due to the challenges of analysing shotgun sequenced metagenomic data such as false negatives/positives and typically low proportion of reads mapped to diet items, shotgun sequencing has not been used for the identification of prey from common vampire bat blood meals. To overcome these challenges and generate longer mitochondrial contigs which could be useful for prey population studies, we shotgun sequenced common vampire bat blood meal samples (n = 8) and utilised a two-step metagenomic approach based on combining existing bioinformatic workflows (alignment and mtDNA contig assembly) to identify prey. After validating our results from detections made through metabarcoding, we accurately identified the common vampire bats' prey in six out of eight samples without any false positives. We also generated prey mitochondrial contig lengths between 138 bp to 3231 bp (median = 770 bp, Q1 = 262 bp, Q3 = 1766 bp). This opens the potential to conduct phylogenetic and phylogeographic monitoring of elusive prey species in future studies, through the analyses of blood meal metagenomic data.

Key Words

animal diet, *Desmodus rotundus*, high-throughput sequencing, molecular ecology, mtDNA assembly, sanguivory, shotgun sequencing

Introduction

The common vampire bat (*Desmodus rotundus*) is one of three extant species of vampire bats (Chiroptera; Phyllostomidae; Desmodontinae) native to Latin America (Greenhall et al. 1983). It has an obligatory sanguivorous diet and feeds on vertebrate blood by biting its prey. The common vampire bat is therefore primed to facilitate cross-species transmission of pathogens such as *Bartonella* (Bai et al. 2012), hemoplasmas (Volokhov et al. 2017), and trypanosomes (Hoare 1965). Common vampire bats are also the primary reservoir of the rabies virus

Copyright *Physilia Y. S. Chua et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

PENSOFT

in much of Latin America (Schneider et al. 2009). Rabies is a lethal zoonotic disease, killing thousands of livestock annually and causing sporadic outbreaks in human populations where bats routinely feed on humans (Schneider et al. 2009). Land-use change from forest to livestock pastures has provided the common vampire bats with an abundant and accessible source of mammalian prey, leading to population growth and range expansion (Delpietro et al. 1992; Lee et al. 2012; Streicker and Allgeier 2016). These bats are also generalist feeders, able to feed not just on livestock but also on wildlife including marine species such as sea lions and penguins (Luna-Jorquera and Culik 1995; Catenazzi and Donnelly 2008). As their distribution continues to respond to climate change, feeding patterns of the common vampire bats can be expected to change accordingly (Hayes and Piaggio 2018). Hence, knowing the diet of the common vampire bat and thereby, their potential routes of disease transmission, is a necessary step to control the spread of pathogens in a cost-effective and efficient manner. One way of determining what taxa they feed from and can potentially transmit pathogens to, is by identifying prey species in vampire bat blood meals (Greenhall 1988; Bohmann et al. 2018).

The different types of methods used for the identification of common vampire bat prey include field observations (Greenhall 1988; Catenazzi and Donnelly 2008), camera traps (Galetti et al. 2016; Calfayan et al. 2018; Zortéa et al. 2018), precipitation tests to visualise antibody-antigen complexes (Greenhall 1970), and stable isotope analysis (Voigt and Kelm 2006; Catenazzi and Donnelly 2008; Streicker and Allgeier 2016). Field observations are challenging as bats are nocturnal (Tournayre et al. 2021), precipitation tests are labour intensive, and stable isotope analysis does not give species resolution (reviewed in Carter et al. 2021). DNA-based methods such as metabarcoding are faster and more precise, allowing for identification of prey down to species level and simultaneous retrieval of common vampire bat population structure (Bohmann et al. 2018). However, amplification of vertebrate prey from blood meals can be challenging due to PCR inhibitors present in blood (Akane et al. 1994), and the co-amplification of vampire bat DNA which could prevent the detection of fragmented prey DNA present in lower copy numbers (Bohmann et al. 2018). To overcome these challenges, blocking primers can be used to reduce the amplification of predator DNA (Vestheim and Jarman 2008; Deagle et al. 2009). However, the design of predator-blocking primers can be made difficult by the lack of DNA reference sequences for prey species, and for the common vampire bat in particular by high intraspecific variation in the common vampire bat mitochondrial genome (Bohmann et al. 2018).

Another DNA-based method that could be used to identify the prey species in the common vampire bat blood meal diet is metagenomics. Metagenomics is where the DNA extracted from samples are shotgun-sequenced without target enrichment of specific markers (Noonan et al. 2005). This has the caveat that sequencing costs are at least ten times more expensive than metabarcoding (Chua et al. 2021). This currently limits the number of samples that can be sequenced using this approach. However, shotgun sequencing overcomes the need to select specific vertebrate primers and can simultaneously retrieve prey and predator mitochondrial DNA, predator gut microbiome and gut parasites (Bon et al. 2012; Paula et al. 2015, 2016; Srivathsan et al. 2015, 2016; Ang et al. 2020). This maximises the amount of information that can be retrieved within one round of sequencing, without the need to carry out additional lab work as is required for metabarcoding using multiple primers.

Despite the advantages of shotgun sequencing, bioinformatics analyses of these metagenomic data can be challenging where false negatives and positives are often an issue (Paula et al. 2016; Chua et al. 2021). Two main strategies can be used to identify metagenomic reads, namely alignment-based and assembly-based approaches. The alignment-based approach is where reads are mapped to a DNA reference database for identification (Zhang et al. 2000), and the resulting mapped reads are identified to the Last Common Ancestor (LCA) for identification (Huson and Weber 2013). However, such identifications are dependent on the completeness of the reference database used (Gómez-Rodríguez et al. 2015; Chua et al. 2021). In diet studies, a low proportion of reads (between 0.0001% and 0.009%) are typically mapped to diet items when using this approach (Srivathsan et al. 2015, 2016; Alberdi et al. 2018; Chua et al. 2021). The reliance on a low proportion of reads to inform results increases the risks of false positives (Paula et al. 2016).

To increase the proportion of informative reads mapped to diet items and generate longer reads, which could be useful for further analysis on prey population structure, an assembly-based approach of assembling mitochondrial DNA (mtDNA) contigs can be carried out using dedicated assemblers (Dierckxsens et al. 2017). However, this is traditionally only used for mtDNA assembly of a single known organism as it requires the selection of an appropriate input reference seed file used for assembly. These reference seed files are usually barcode markers of the known organisms which are used for extending the reads to generate longer mtDNA contigs. In known-mixed template samples, assemblies are limited to the taxa with the highest proportion of reads in the sample (reviewed in (Sharpton 2014)). The assembly of low abundance sequences will be fragmented if the sequencing depth is too low. This can be problematic in diet studies given that predator sequences would overwhelm the proportion of prey sequences. Conserved regions shared between prey and predator could also lead to the assembly of chimeric sequences (Bon et al. 2012).

In animal dietary studies, shotgun sequencing has been used to identify plants in the diets of herbivores (Srivathsan et al. 2015, 2016; Chua et al. 2021), arthropod prey from arthropod predators (Paula et al. 2015, 2016), and vertebrate prey from vertebrate predators (Bon et al. 2012). In common vampire bat studies, shotgun sequencing of

faeces, saliva, or rectal swabs has been used to identify the bat's genomic adaptations to sanguivory (Mendoza et al. 2018), profile its gut microbiome (Mendoza et al. 2018), characterise its viral communities (Bergner et al. 2019; Bergner et al. 2020a), and to assemble the genomes of common vampire bat viruses (Bergner et al. 2020b). However, identification of prey from common vampire bat blood meal samples has not been carried out using the metagenomics shotgun sequencing approach.

Here, we use shotgun sequencing for the first time to identify prey from common vampire bat blood meal samples. To overcome the challenges associated with each of the shotgun sequencing approaches outlined above, we demonstrate the advantages of combining both the alignment and assembly-based approaches in a stepwise manner to reduce the limitations associated with each approach. Additionally, the inclusion of the second assembly-based approach acts as a proof-of-concept that we can retrieve longer mitochondrial contigs of vampire bats' prey from their blood meals, which could be useful for future prey population genetics studies. To validate our results, we verified our metagenomic prey identification with the metabarcoding results obtained from the same samples by Bohmann et al. (2018).

Materials and methods

Sample information

The eight analysed blood meal samples from common vampire bats (*Desmodus rotundus*) were collected in Peru between 2009 and 2013 at four sites across two ecoregions; Amazon (MDD130) and Pacific coast (LMA4, LMA6, and LMA10) (Suppl. material 1: Table S1). Four samples were collected each from the Amazon (sample 25, 54, 116, and 121) and the coastal ecoregions (sample 29, 70, 90, and 94). The blood meal collection followed the procedures outlined in Streicker and Allgeier (2016). Using mist nets and/or harp traps for the capture of bats outside daytime roosting sites, captured bats were first anaesthetised with ketamine (8.3 to 12.5 mg/kg) before 50 µL of blood per bat were extracted from the stomach with an empty syringe attached to a sterile 5-French nasogastric feeding tube (Bohmann et al. 2018).

From the metabarcoding results outlined in Bohmann et al. (2018), the prey taxa of the eight common vampire bat blood meal samples were identified as chicken (*Gallus gallus*), cow (*Bos taurus*), donkey (*Equus asinus*), horse (*Equus caballus*), pig (*Sus scrofa*), and the South American tapir (*Tapirus terrestris*) (Fig. 1).

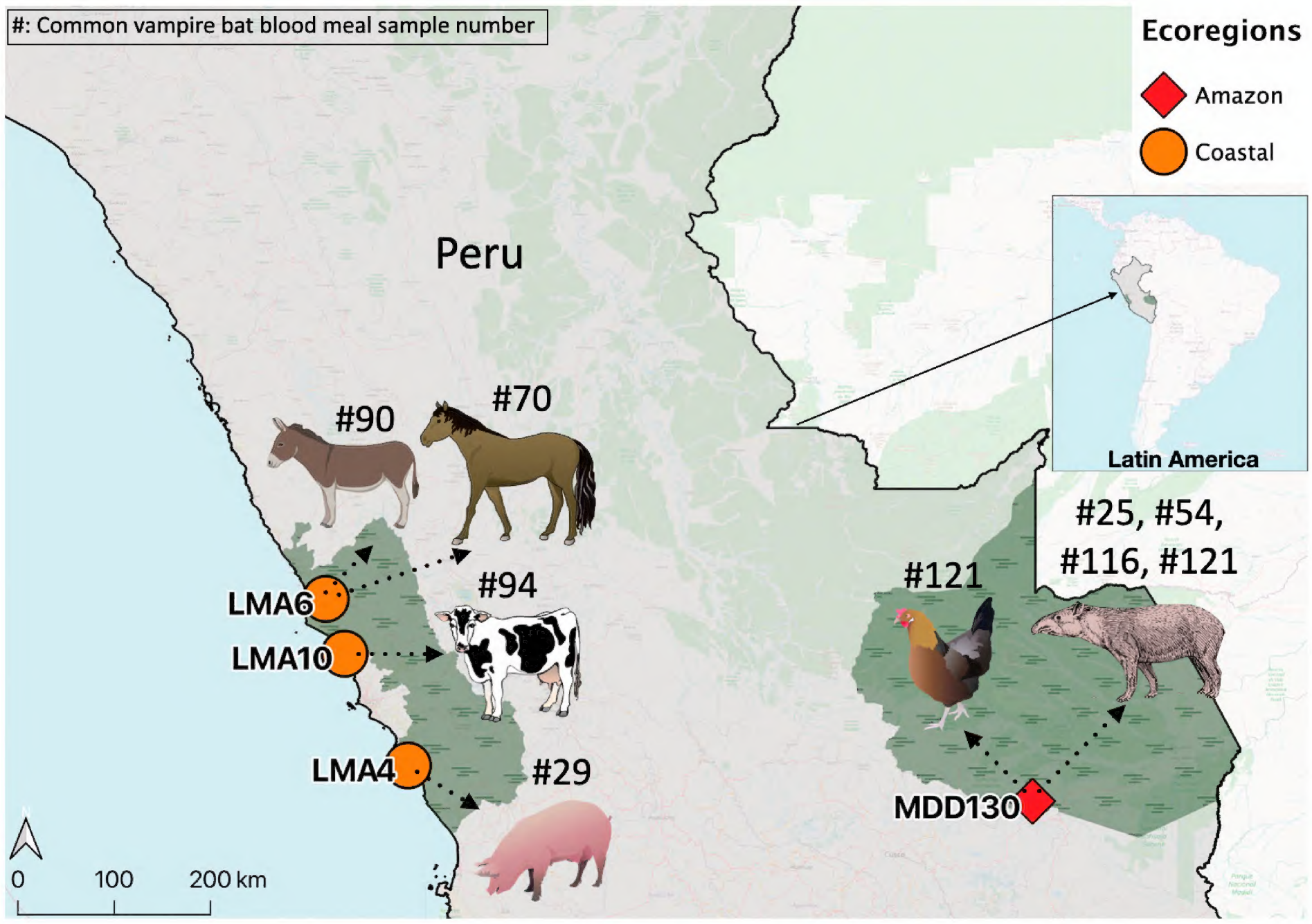

Figure 1. Collection sites of eight common vampire bat (*Desmodus rotundus*) blood meal samples from two ecoregions in Peru, showing the identity of the prey taxa derived from metabarcoding analysis (Bohmann et al. 2018). Map created in QGIS version 3.12. Some of the elements included in the figures were obtained and modified from the Integration and Application Network, University of Maryland – Center for Environmental Science (https://ian.umces.edu/symbols/), and BioRender.com. Image of tapir from Foresman (2007).

Metagenomics laboratory workflow

We used the extracted DNA from the eight blood meal samples that were previously extracted as described in Bohmann et al. (2018). DNA extraction was carried out using the QIAGEN Investigator Kit, following the manufacturer's instructions (Isolation of DNA from FTA and GUTHRIE cards, version 2). DNA extracts were fragmented using a Diagenode Bioruptor (Diagenode) using a program of eight cycles with 15 seconds on and 90 seconds off, targeting a fragment size of 500 bp. 32 µL of fragmented DNA was used to generate Illumina shotgun sequencing libraries using the blunt-end single-tube library preparation protocol (Carøe et al. 2018) with modifications from Mak et al. (2017). The libraries were purified using SPRI bead purification according to (Rohland and Reich 2012). Specifically, 100 µL of bead solution was added to each library (60 µL), incubated for 5 minutes, washed twice in 80% ethanol and eluted in 30 µL of 10 mM Tris-HCl by heating to 40 °C for 5 minutes. Libraries were evaluated with quantitative PCR (qPCR), 480 Lightcycler 2× qPCR mastermix (Roche) in 10 µL reactions, with 0.2 µM primer IS7 and IS8 (Meyer and Kircher 2010), and 1 µL of 10× diluted library. Based on cyclic threshold values, libraries were given 7 to 11 PCR cycles for index PCR, using full-length Illumina primers with indexed P7 adapters. This was done using 10 µL library in a 50 µL PCR reaction consisting of 0.25 mM dNTP (Invitrogen), 0.2 mM forward and reverse primer, 0.1 U/mL Taq Gold polymerase (Applied Biosystems), 1× Taq Gold buffer (Applied Biosystems), 2.5 mM $MgCl_2$ (Applied Biosystems), and 0.8 mg/mL BSA (New England Biolabs). PCR consisted of 10 minutes denaturation and activation at 95 °C, followed by 7 to 11 cycles of 30 seconds at 95 °C, 30 seconds at 60 °C, and 1 minute at 72 °C, followed by a final extension at 72 °C for 5 minutes before cooling to 4 °C. Libraries were purified using MinElute spin columns (Qiagen). Purified libraries were quantified on an Agilent 2100 bioanalyzer (Agilent technologies) before equimolar pooling. Sequencing was carried out on one lane of an Illumina 2500 Hiseq instrument (Illumina Inc.) using 125 cycle chemistry in paired-end (PE) mode at the GeoGenetics Sequencing core, University of Copenhagen, Denmark.

Metagenomics bioinformatics workflow

Between ~17 and ~36 million paired-end (PE) reads were generated per blood meal sample. Adapter removal, quality trimming of sequences with Phred quality score less than 30, and removal of reads shorter than 85 base pairs (bp) were carried out with Trim galore v0.5.0 (Andrews et al. 2015). This cut-off length at two-thirds of the sequenced reads was introduced to reduce the rates of false-positive identification in downstream analysis, while keeping most true-positive reads (Srivathsan et al. 2015, 2016; Chua et al. 2021). FastQC v0.11.9 was used for quality checks before and after filtering (Andrews 2010). For each sequenced blood meal sample, we used the Burrows-Wheeler Alignment software with the Maximal Exact Matches algorithm v0.7.17 (BWA-MEM) and Sequence Alignment/Map v1.9 (SAMtools) software to map and align PE reads to the common vampire bat genome downloaded from NCBI GenBank (RefSeq assembly accession: GCF_002940915.1) (downloaded 27.04.20). Aligned reads mapped to the common vampire bat genome were subsequently removed from each of the sequenced blood meal samples (Li et al. 2009; Li 2013). Using the Browser Extensible Data v2.29 (BEDtools) software (Quinlan and Hall 2010), we converted the BAM files generated in the common vampire bat sequence-removal step to fastq files, using the bamToFastq function, for downstream analysis.

Prey identification

The analyses of metagenomic reads were carried out without any prior knowledge and access to the corresponding metabarcoding data of the same samples. Hence, the metagenomic analyses were conducted in a 'blind' manner, where the metabarcoding data was only accessed after analyses had been concluded to validate the metagenomics results.

For the identification of common vampire bat blood meal prey taxa, we used a two-step approach. In the first step, we used an alignment method, in which we carried out the Basic Local Alignment Search Tool (BLAST) to map reads to reference data. For the second step, based on the BLAST results, we used a seed (COI barcode or whole mtDNA) corresponding to the identity of the prey to carry out the assembly of prey mtDNA contigs. The mtDNA contig assembly step was also used to retrieve any additional identification of prey from blood meal samples that had no results from the BLAST-alignment step.

Step 1: BLAST-alignment step

We generated a reference database by downloading taxonomically informative barcodes consisting of metazoan mitochondrial cytochrome oxidase subunit 1 (COI) sequences from the Barcode of Life Data System v4 (BOLD) (473,748 sequences forming 38,618 BINS, representing 33,299 metazoan species downloaded 27.04.20) (Ratnasingham and Hebert 2007). MEGABLAST searches for the sequenced bloodmeal PE reads were conducted against the generated COI barcode reference database (word size = 28, percentage identity = 98%) (Camacho et al. 2009; Srivathsan et al. 2015, 2016). We used the bold package in R with a custom R script BOLD_taxID to retrieve taxonomy classification details for each BOLD BIN in the generated COI barcode reference database (retrieved 29.04.20). For taxonomic assignment of the common vampire bat blood meal sequences to determine prey species, we used a custom R script BOLD_readsidentifier with the following filtering parameters of 98% sequence identity and 85 bp overlap of a given read with the COI

barcode (Srivathsan et al. 2015, 2016). This threshold was determined from other metagenomic studies showing that a minimum of 98% sequence identity and at least two-thirds length overlap of a given read is required to eliminate false assignments (Paula et al. 2015, 2016; Srivathsan et al. 2015, 2016; Chua et al. 2021). Only sequences identified as belonging to the classes Aves or Mammalia were kept. Following the Lowest Common Ancestor algorithm (Huson and Weber 2013), we obtained species-level identification for a given read if the BLAST hit was to one species, genus-level if it matches to two or more species from one genus, and family level if it matches to two or more genera from the same family. We only retained identifications at a given taxonomic hierarchy where there were no conflicts in identification between the forward and reverse reads (Srivathsan et al. 2015, 2016). For samples with more than one species identified, we only used the identity of the species with the highest number of reads matched. This is supported by the expectation that vampire bats feed on a single individual per night, thus secondary reads are likely to represent false positives (Greenhall 1988).

Step 2: mtDNA contig assembly step

For the mtDNA contig assembly step, we used the organelle assembler and heteroplasmy caller software NOVOPlasty v3.8.3 (Dierckxsens et al. 2017) to attempt assemblies of prey mtDNA contigs. NOVOPlasty is a seed-based assembler, where assembly is initiated by a reference seed file. Reference seed files of specific prey species identified from each blood meal sample from the initial BLAST-alignment step were used as inputs for mtDNA contig assembly. If more than one prey species was identified per blood meal sample from the BLAST-alignment step, we used reference seed files of the species with the highest number of reads mapped to the COI barcode. For blood meal samples with no prey identified from BLAST, we used reference seed files of all prey species determined from the first BLAST-alignment step. Reference seed files used were either species-specific COI barcodes retrieved from BOLD (downloaded 17.06.20) or whole mtDNA retrieved from NCBI GenBank (downloaded 27.08.20) (Suppl. material 1: Tables S2–S4). We carried out assembly with a K-mer size of 39 starting with a species-specific COI barcode as seed, and decreasing the K-mer size to 27 if no contigs were assembled. We changed the input seed files to species-specific mtDNA if no contigs were assembled after decreasing K-mer size to 27. Assembled mtDNA contigs were checked by using the web BLAST blastn suite, to obtain the closest match (Madden 2013). Contigs were also manually checked for alignment to reference seed using Geneious Prime v2020.2 (https://www.geneious.com). To ensure the accuracy of the mtDNA contig assembly step, we attempted mtDNA contig assembly with seed files of species not identified as prey from the BLAST-alignment step for each corresponding blood meal sample with K-mer size 39 (COI barcode of all species in Suppl. material 1: Table S3, and mtDNA of *Tapirus terrestris*). Any contig(s) assembled were checked by uploading the contig sequences to the web BLAST blastn suite for sequence identity (Suppl. material 1: Table S5). Only the closest matched species in terms of query cover and percent identity (> 90%) were kept.

After these two steps, the metagenomics outputs from this study were compared with the metabarcoding results from Bohmann et al. (2018) to check for any discrepancies between the two approaches. After validation of the metagenomic data with metabarcoding results, we carried out an additional analysis using the mtDNA contig assembly step to determine if we could retrieve dietary information and mitochondrial contigs from more samples with preliminary knowledge of diet taxa. This step was carried out to determine if metagenomics could potentially be used for future in-depth analysis of prey population genetic studies if the diet were known or previously informed by other methods such as metabarcoding. For this diet-informed additional analysis, reference seed files of all prey identified from the eight blood meal samples through metabarcoding were used (*Gallus gallus*, *Bos taurus*, *Equus asinus*, *Equus caballus*, *Sus scrofa* and *Tapirus terrestris*) (Suppl. material 1: Tables S3–S5). Of these seed files, only *Sus scrofa* had not been identified as a possible prey from the initial 'blind' metagenomic analyses.

Results

Reads mapped to the common vampire bat genome (*Desmodus rotundus*) made up 48.1% to 98.7% of PE reads generated per blood meal sample (~16 million to ~60 million reads, median 50 million reads, Q1 = 20 million reads, Q3 = 55 million reads). After removal of common vampire bat sequences, 1.3% to 51.9% of reads remained which includes sequences from the prey, the common vampire bat's gut microbiome, and gut parasites (~67,000 to ~17 million reads per sample, median 5.7 million reads, Q1 = 1.2 million reads, Q3 = 10 million reads) (Table 1).

Prey identification

In the BLAST-alignment step, BLAST searches against the COI database yielded between 1 and 36 reads mapped to Mammalia or Aves (<0.0001% – 0.003% of vampire-bat removed reads, <0.0001% – 0.0001% of total sequenced reads) for five of the blood meal samples (samples 54, 70, 90, 94 & 121). Three blood meal samples did not have any reads mapped to the COI barcode for either Mammalia or Aves (samples 25, 29, & 116) (Suppl. material 1: Table S6). These three bloodmeal samples each contained more than 95% of common vampire bat sequences prior to common vampire bat sequence removal. For the five blood meal samples with reads mapped to the COI barcode, the majority of the reads were assigned species-level identification (57.6%), 37.3% of the reads

were assigned to genus, 3.4% of the reads assigned to family, and 1.7% of the reads assigned to order. For these five blood meal samples, four had reads assigned to only one species except for sample 94 with five species identified (1 read: *Bison bonasus*, 1 read: *Bos grunniens*, 2 reads: *Bos indicus*, 1 read: *Bos primigenius*, 15 reads: *Bos taurus*). We only kept the identity of the species with the highest number of reads matched, which was *Bos taurus*. The species identity of the prey found in the five blood meal samples were; *Tapirus terrestris* (sample 54), *Equus caballus* (sample 70), *Equus asinus* (sample 90), *Bos taurus* (sample 94), and *Gallus gallus* (sample 121).

For the assembly of mtDNA contigs, the proportion of reads assembled was between 0.0006% (10 reads) to 0.2% (3500 reads) of vampire-bat removed reads (Table 1). Based on the initial metagenomics results without the analyses carried out in the diet-informed additional step, the median length of mtDNA contigs assembled was 770 bp (Q1 = 262 bp, Q3 = 1766 bp). The smallest contig was 138 bp from sample 121 and the largest was 3231 bp from sample 94 (Table 2). After the inclusion of the diet-informed additional analysis carried out, the median length of mtDNA contigs assembled was 985 bp (Q1 = 256 bp, Q3 = 1152 bp). Most samples had only one contig assembled except for samples 54 and 94, with two and three contigs assembled respectively. The species identity of the prey from the assembled mtDNA contigs corresponded to the BLAST-alignment results for all five blood meal samples (samples 54, 70, 90, 94 & 121). We retrieved prey information for sample 25 from the mtDNA contig assembly which had no identification after the first BLAST-alignment step. The identity of the prey from this blood meal sample was *Tapirus terrestris*. After the inclusion of the diet-informed additional mtDNA contig assembly step, we were able to retrieve prey identification for sample 29 (*Sus scrofa*) which did not have any results using the analyses conducted without any prior dietary information. All identified prey were identified to the species level. We did not retrieve any prey identification for sample 116 with our metagenomics workflow even with the additional mtDNA contig assembly step informed by metabarcoding results (Fig. 2).

When comparing the metagenomics outputs with the metabarcoding analysis presented in Bohmann et al. (2018), there was a 100% congruence in the identity of the prey detected using metagenomics for all samples, with no false positives. However, metagenomics failed to identify *Tapirus terrestris* found in the blood meals of samples 116 and 121, as well as *Sus scrofa* found in sample 29, leading to a 33% false negative detection rate (Fig. 3). However, with access to previous metabarcoding data from the same samples, the diet-informed analysis additionally identified *Sus scrofa* in sample 29, reducing the false negative detection rate to 22%.

Table 1. Proportion of metagenomic reads belonging to the common vampire bat (*Desmodus rotundus*) and proportion of reads assembled (common vampire bat-removed sequences) from the mtDNA contig assembly step for prey identification without prior dietary information. *Prey identification was only included after additional analyses were carried out with access to dietary information as determined from a previous metabarcoding study of the same common vampire bat blood meal samples (Bohmann et al. 2018).

Samples	Total reads	Bat sequences	Proportion of bat sequences	Bat-removed reads for downstream analysis	Assembled mtDNA reads	Proportion of reads assembled for prey identification
25	57782738	55744368	96.5%	2038370	40	0.00196%
*29	61585369	60726115	98.6%	859254	50	0.00582%
54	30704614	21223068	69.1%	9481546	3478	0.03668%
70	61417998	51802288	84.3%	9615710	48	0.00050%
90	56874109	55157869	97.0%	1716240	3500	0.20393%
94	31315790	19574502	62.5%	11741288	3010	0.02564%
116	50267627	49596878	98.7%	670749	0	0.00000%
121	33684838	16217550	48.1%	17467288	10	0.00006%

Table 2. Prey identity of assembled mtDNA contigs for each common vampire bat (*Desmodus rotundus*) blood meal shotgun-sequenced sample with query cover and percent identity score. *Prey identification was only included after additional analyses were carried out with access to dietary information as determined from a previous metabarcoding study of the same common vampire bat blood meal samples (Bohmann et al. 2018).

Sample	Contig length (Bp)	Closest Match	Query Cover	Percent identity	Accession number
25	684	*Tapirus terrestris*	100%	99.42%	AJ428947.1
*29	166	*Sus scrofa*	92%	98.70%	MK251046.1
54	1152	*Tapirus terrestris*	100%	99.39%	NC_053962.1
	770	*Tapirus terrestris*	97%	99.74%	NC_053962.1
70	268	*Equus caballus*	98%	100%	MN187574
90	256	*Equus asinus*	100%	100%	MK982180.1
94	3231	*Bos taurus*	100%	100%	MN714183.1
	2380	*Bos taurus*	100%	100%	AY676863.1
	802	*Bos taurus*	100%	100%	MZ901759.1
121	138	*Gallus gallus*	100%	100%	LC082227.1

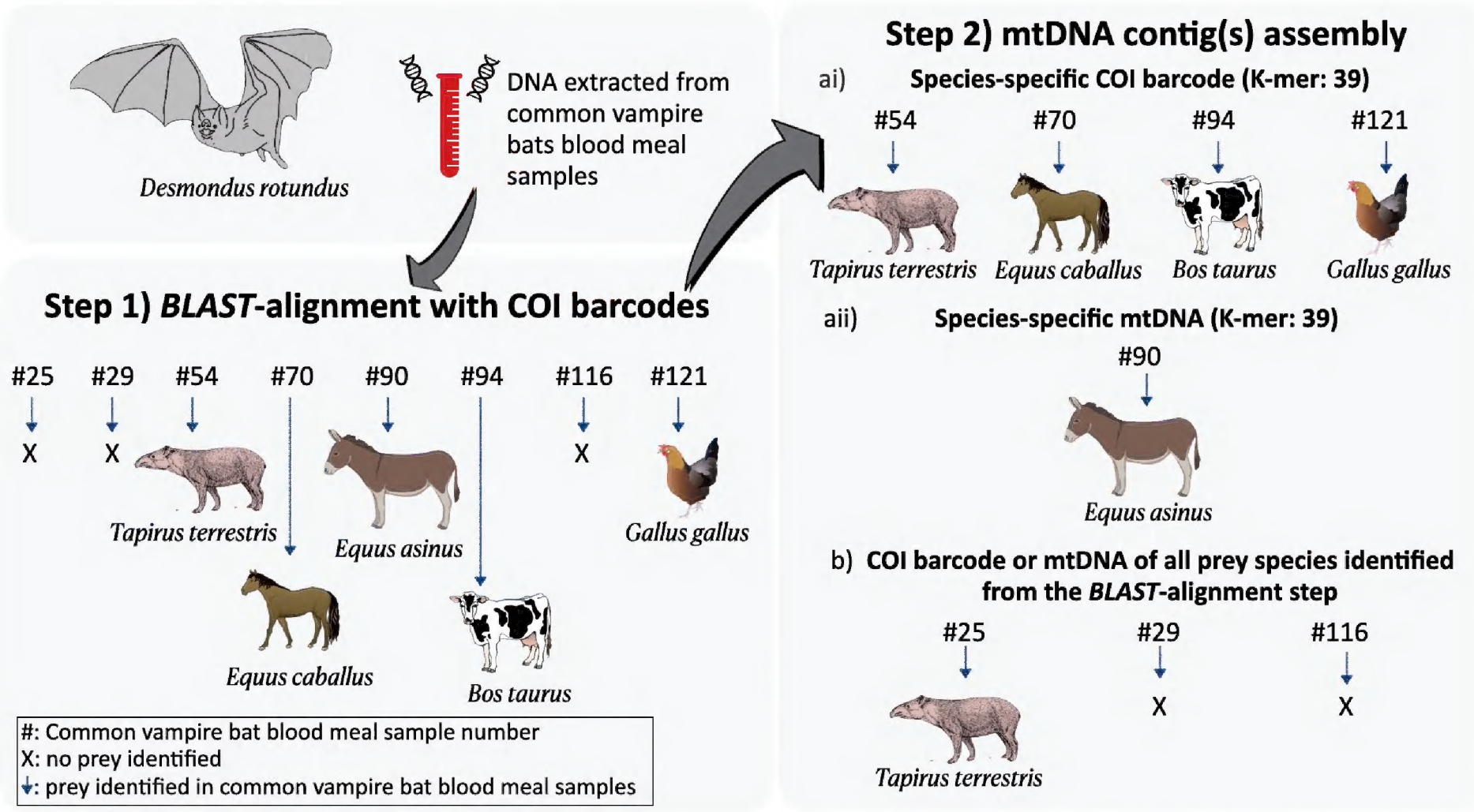
Desmondus rotundus
DNA extracted from common vampire bats blood meal samples
Step 1) BLAST-alignment with COI barcodes
#25 #29 #54 #70 #90 #94 #116 #121
X X X
Tapirus terrestris
Equus asinus
Gallus gallus
Equus caballus
Bos taurus
#: Common vampire bat blood meal sample number
X: no prey identified
: prey identified in common vampire bat blood meal samples
Step 2) mtDNA contig(s) assembly
ai) Species-specific COI barcode (K-mer: 39)
#54 #70 #94 #121
Tapirus terrestris
Equus caballus
Bos taurus
Gallus gallus
aii) Species-specific mtDNA (K-mer: 39)
#90
Equus asinus
b) COI barcode or mtDNA of all prey species identified from the BLAST-alignment step
#25 #29 #116
X X
Tapirus terrestris

Figure 2. Prey identification of common vampire bat (*Desmodus rotundus*) blood meal shotgun-sequenced samples using a two-step metagenomics approach. In the first BLAST-alignment step, metagenomic reads were mapped to COI barcodes using BLAST. Second, mtDNA assembly of prey contigs was carried out with Novoplasty using ai) seeds from COI barcode or aii) mitochondrial DNA (mtDNA) of prey identified from the BLAST-alignment step corresponding to each sample, and b) seeds from COI barcode or mtDNA of all prey species identified from the BLAST-alignment step. Some of the elements included in the figures were obtained and modified from the Integration and Application Network, University of Maryland – Center for Environmental Science (https://ian.umces.edu/symbols/), and BioRender.com. Image of tapir from Foresman (2007). Common vampire bat image credit: Megan Griffiths.

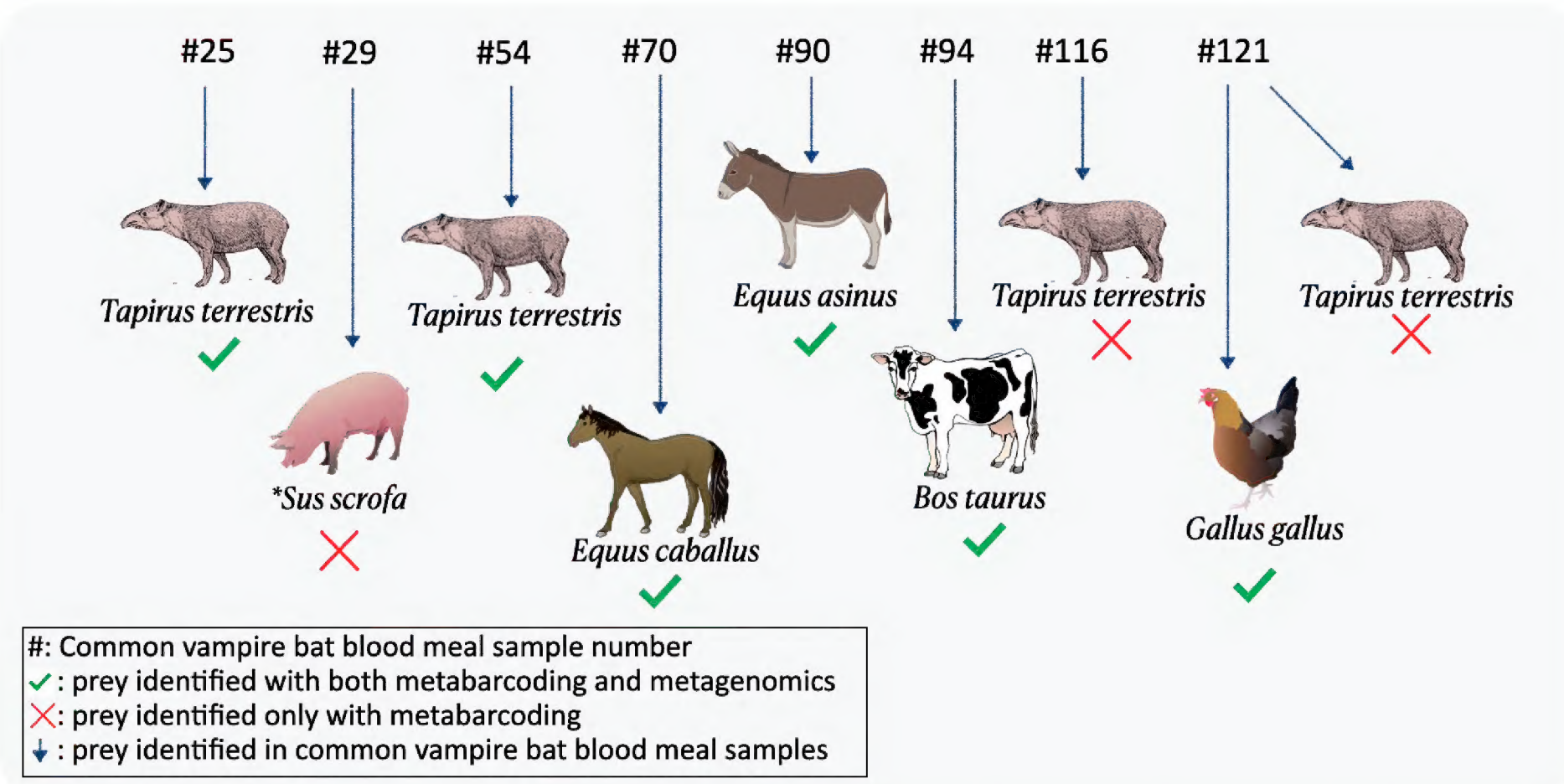
#25 #29 #54 #70 #90 #94 #116 #121
Tapirus terrestris
Tapirus terrestris
Equus asinus
Tapirus terrestris
Tapirus terrestris
*Sus scrofa
Equus caballus
Bos taurus
Gallus gallus
#: Common vampire bat blood meal sample number
: prey identified with both metabarcoding and metagenomics
: prey identified only with metabarcoding
: prey identified in common vampire bat blood meal samples

Figure 3. Prey identification of common vampire bat blood meal samples using shotgun metagenomics as compared to metabarcoding (Bohmann et al. 2018). Green ticked symbols signify consensus between both high-throughput sequencing (HTS) approaches. Red crossed symbols signify prey taxa identified using metabarcoding but not identified using metagenomics. **Sus scrofa* was identified in sample 29 only after additional analyses were carried out with access to dietary information as determined from a previous metabarcoding study of the same common vampire bat blood meal samples (Bohmann et al. 2018). Some of the elements included in the figures were obtained and modified from the Integration and Application Network, University of Maryland – Center for Environmental Science (https://ian.umces.edu/symbols/), and BioRender.com. Image of tapir from Foresman (2007).

Discussion

In this study, we demonstrated how metagenomics can be used to identify common vampire bat prey in blood meals. We utilised a two-step strategy, combining both the alignment and assembly approaches to obtain prey identification. These metagenomic prey identifications were subsequently compared to previous metabarcoding results for validation (Bohmann et al. 2018). The metabarcoding results were also used in the additional diet-informed analysis to see if we could obtain any additional diet information from samples that had no prey identified from our initial 'blind' metagenomics workflow.

In the BLAST-alignment approach, the proportion of reads mapped to diet items (<0.0001% – 0.003%) is similar to other metagenomic studies (Srivathsan et al. 2015, 2016; Alberdi et al. 2018; Chua et al. 2021). Despite the low proportion of reads mapped, we achieved good species resolution at 57.6%, which is higher than these recent metagenomic diet studies. However, in samples with a high proportion of vampire bat DNA (> 95% in samples 25, 29, and 116), we were unable to map any reads to prey. This could be due to the extremely low copy numbers and fragmented nature of prey DNA present. The completeness of the reference database used for the matching of short reads could also have resulted in these missing identifications (Chua et al. 2021). Without carrying out the second step mtDNA contig assembly approach, stopping at this step of the bioinformatics analysis would result in missing prey identification for one sample (sample 25). Through our stringent BLAST-alignment parameters, we were able to minimise false-positives at this step. This reduced the number of possible reference seed files used as input for the second mtDNA contig assembly step and decreased the computational effort required for carrying out contig assembly with multiple seed files for a given sample. To better document the extent of spurious matches at less stringent filtering parameters, we recommend that future studies explore their data using different filtering parameters, albeit with an increase in computational effort. Our preliminary tests using a filtering parameter of 98% sequence identity but with only 50% of reads overlap (63 bp) (Suppl. material 1: Table S7) showed some spurious hits to other classes such as Actinopterygii, Amphibia, and Ascidiacea. However, we did not explore the effects of reduced sequence identity, which could be investigated further in future studies. A recommendation would be to carry out *in silico* analysis to determine what will be the most optimal filtering parameters to use for any given dataset (Chua et al. 2021). Even though not observed in our preliminary tests, less stringent filtering parameters might lead to better retrieval of diet items, reducing the risk of false negatives. We also expect that any false positives arising from utilising less stringent filtering parameters would likely not lead to the assembly of large contigs in the next step.

From the mtDNA contig assembly approach, a higher proportion of reads were assembled as compared to the BLAST-alignment step (0.0006% – 0.2%). This is expected because reads were only mapped to the COI barcode in the BLAST-alignment step while in the mtDNA contig assembly step, longer mtDNA contigs were generated. Even though the current standard approach used in existing metagenomic dietary studies utilises only the BLAST-alignment step (Paula et al. 2015, 2016; Srivathsan et al. 2015, 2016; Alberdi et al. 2018; Chua et al. 2021), the inclusion of the second step mtDNA contig assembly confers several advantages over just utilising the BLAST-alignment step alone for the identification of diet items. The first is that the resolution of prey identified was at 100% species-level as compared to just 57.6% using the BLAST-alignment step. Second, including this approach in the metagenomics workflow resulted in an 11% increase in samples with diet identified. Finally, longer fragments generated by this second step are more informative than shorter ones retrieved from the first BLAST-alignment step. For example, longer mitochondrial fragments can be used for phylogenetic analysis of rare or elusive prey as demonstrated by Nguyen et al. (2021) through metabarcoding of leech blood meals. As common vampire bats can feed on species of conservation interest such as the American Tapir (*Tapirus terrestris*) (Flesher and Medici 2022), metagenomic analyses of the common vampire bat blood meals can act as a proxy to monitoring these species that are elusive and challenging to monitor. With larger sample sizes, the inclusion of this second step would not only provide missing identification for more samples with greater species resolution, but more importantly, it can generate longer mitochondrial contigs of diet items that could be useful for other types of applications such as phylogenetic and phylogeographic studies. Additionally, good quality assemblies from blood meals can potentially be used as DNA reference data, offering an alternative source of sampling species for genetic materials (Wilting et al. 2021). Given that false-positives remains one of the main challenges of analysing metagenomic data, longer fragments give a better confidence of the species identified as compared to using only the first BLAST-alignment step of matching short barcodes. Even though the percent identity of contigs in relation to web BLAST results were at least 98.7%, diet species that are not represented in database could lead to inaccurate identifications in future studies. Increasing the sequencing depth could help in reducing fragmented assemblies caused by low abundance sequences (Sharpton 2014).

Even though the application of the mtDNA contig assembly step for eukaryote mtDNA assembly can be problematic due to complex genomes and low abundance (Azam and Malfatti 2007), prior to the diet-informed analysis carried out, we managed to retrieve prey identification for most samples except for two (samples 29 and 116), which had the highest proportion of common vampire bat sequences at 98.6% and 98.7% respectively.

The high abundance of common vampire bat sequences found in blood meals could have drowned out the DNA sequences belonging to the prey during sequencing. This feeds into the limitation of the mtDNA contig assembly step, where assemblies are limited to taxa with a high proportion of reads in a sample (Sharpton 2014). Amongst all eight samples, sample 121 was the only blood meal sample found to have had two prey identified using metabarcoding (Bohmann et al. 2018). This could also be the reason why we were only able to detect one prey, *Gallus gallus* and not *Tapirus terrestris*, in the blood meal of sample 121. It could also have led to the assembly of a short 138 bp *Gallus gallus* contig which would not have been useful for phylogenetic analyses, further highlighting the issues of this second mtDNA contig assembly step in dealing with complex diets. To assemble longer contig lengths that could be useful for downstream phylogenetic analyses, two issues must be considered. The first is the complexity of the host's diet where only the most abundant diet item can be assembled as shown in our study. The second is access to fresh blood meal samples as short DNA fragments persist for a much longer time as compared to long fragments (Deagle et al. 2006; Reeves et al. 2018), possibly leading to fragmented assemblies.

Another concern was the selection of appropriate reference seed files for mtDNA contig assembly. The selection of inappropriate reference seed could lead to Type I errors resulting in false positives and inaccurate determination of diet. However, when we tested for the accuracy of mtDNA assembly by using reference seed files from other species, eg: *Equus asinus* for *Tapirus terrestris* in sample 54, the assembler chosen in our study managed to accurately assemble mtDNA contigs belonging to the identified prey. Based on our samples and the assembler used, the selection of input reference seed files did not result in any false positives but further tests should be carried out on larger sample sizes with a variety of reference seed files. It would also be interesting to test if reference seed files of different genetic regions from distantly related species would have any effect on assembly accuracy. Until further tests can be carried out on whether the choice of reference seed files could affect assembly, researchers utilising such assemblers should first carry out a BLAST-alignment step to inform them on the choice of reference seed files to use to optimise computational effort and prevent inaccurate identifications. Another suggestion to improve on the mtDNA assembly would be for future studies to tests if k-mer based *de novo* assemblers such as Norgal can be used for complex metagenomic reads, as such assemblers overcome the need to select appropriate input reference seed files (Al-Nakeeb et al. 2017). Even though this approach is commonly used for organisms with small genomes such as bacteria (Forouzan et al. 2017), *de novo* assembly of mammalian genomes from metagenomic data is more challenging due to the larger, more complex genomes. As such, seed-based assembly like NOVOPlasty is preferred for the assembly of more complex mammalian genomes.

When we compared our results to previous metabarcoding analyses carried out on the same samples (Bohmann et al. 2018), we were able to accurately identify common vampire bat blood meal prey from six of the eight samples. Our diet-informed analysis of using reference seed files of all prey identified with metabarcoding filled the missing diet information for sample 29. However, we recognise that it is not feasible to carry out both metabarcoding and metagenomics just for diet identification due to the extra costs and workload. Hence, the drawback of using metagenomics to assess the common vampire bat diet is that not all prey can be identified, and sequencing costs are at least ten times more than for metabarcoding (Chua et al. 2021). Given that metabarcoding is cheaper, it will remain the go-to technique when it comes to molecular diet profiling of animals. The current costs of shotgun sequencing mean that published metagenomic animal dietary studies are typically limited to only a few individuals (Bon et al. 2012; Srivathsan et al. 2015, 2016; Paula et al. 2016; Alberdi et al. 2018; Chua et al. 2021). As such, bioinformatics procedures are still in their infancy for analysing metagenomic dietary datasets and pipelines used differ based on the type of diet, and the taxonomic class of the predator in comparison to its diet. Our strategy based on combining two existing bioinformatic workflows; the alignment and assembly-based approaches, can help to advance metagenomic dietary research and open doors for further applications including the phylogenetic analyses of diet taxa. The strategy outlined here is useful in scenarios where additional information about the prey is required, for example in population genetic studies, with the caveat that this two-step approach could be more computationally intensive and additional streamlining is needed to optimise performance. Nevertheless, this two-step metagenomics approach based on our small sample size showed that there were no false-positives, which is an important challenge to overcome when working with metagenomic datasets. For this approach to be robust for various types of datasets, it should be tested on larger sample sizes or even through *in silico* means to get a better overview of the false-positive and false-negative rates that can be expected with metagenomic data.

Future outlook

We demonstrated here as a proof-of-concept, how longer mtDNA contigs of diet items can be generated from metagenomic data sequenced from common vampire bat blood meal samples. This offers great potential to other applications such as the phylogenetic analyses of diet items and can act as a proxy to monitoring rare and elusive prey. Through metagenomics, blood meals can also provide an alternative source of sampling species for creating DNA reference data. Our metagenomics approach can be used to cherry-pick samples that have been previously analysed using metabarcoding to zoom in on specific prey populations for such applications. Additionally, existing metagenomic

data that have not been analysed for diet can also be repurposed and analysed without the need to carry out additional sequencing. For example, metagenomic data that were sequenced for virus discovery or parasite detection provides an opportunistic source of data that could also be analysed for host diet (Bergner et al. 2021a, 2021b). To fully utilise the potential of metagenomics, improvements in bioinformatic procedures are still required to optimise data analyses and to make the most of all sequenced data such as streamlining bioinformatics pipelines to reduce computational requirements. As large amounts of data are generated through metagenomics, the use of this method for the sole purpose of diet identification alone would be wasteful, whereas metabarcoding would be a cheaper alternative. Hence, we envision that metagenomics would be more useful when the research question also targets other types of information such as host population studies, gut microbiome composition, and gut parasites which can be analysed from the same metagenomic dataset (Srivathsan et al. 2015, 2016; Mendoza et al. 2018; Bergner et al. 2021a). For example in our study, a large proportion of metagenomic reads are mapped to the predator itself (48.1% to 98.7%) which could be used for population genetic studies of the common vampire bat. Future work should also explore how the metagenomics approach outlined in our study performs for the assembly of vertebrate mtDNA in more complex sample types such as faeces or bulk invertebrate samples. As sequencing costs are decreasing, more computationally efficient bioinformatics pipelines can be developed to overcome some of the challenges mentioned. This can lead to new ways of analysing metagenomic data, revealing the potential of utilising this technique for a wide array of molecular applications including phylogenetics and phylogeographic surveillance or even the contribution of DNA reference data. With this, we can expect there to be a shift towards more metagenomic studies in the coming future.

Author contributions

PYSC: Conceived the study, carried out bioinformatics analysis, data analysis and visualisation, and wrote the manuscript; CC: Carried out laboratory analysis and contributed to the writing of the manuscript; AP: Wrote the custom R scripts used in the BLAST-alignment approach and supervised data analysis; CSRA: carried out the removal of vampire-bat sequences and contributed to the writing of the manuscript; GJ: Contributed to the writing of the manuscript; DS: Collected samples and contributed to the writing of the manuscript; KB: Conceived and supervised the study, carried out laboratory analysis, and contributed to the writing of the manuscript.

Data accessibility

Sequencing data and FastQC reports are available at the University of Copenhagen Electronic Research Data Archive (UCPH ERDA) Digital Repository, DOI: https://doi.org/10.17894/ucph.5601175c-e786-43ef-ab40-15026a16303a. Assembled contigs, NOVOPlasty log files, and scripts used for data generation is available at figshare: https://doi.org/10.6084/m9.figshare.19154321.v1.

Competing interests

The authors have no competing interests.

Acknowledgements

We thank the staff at GeoGenetics Sequencing core, University of Copenhagen, Denmark, for sequencing. PYSC was supported by the European Union Horizon 2020 research and innovation programme under grant agreement No 765000, H2020 MSCA-ITN-ETN Plant.ID network. CC was funded by the Carlsberg Foundation (CF18-1110; "Archives") Fieldwork was supported by the US National Science Foundation (DEB-1020966). DS was funded by a Wellcome Senior Research Fellowship (217221/Z/ 19/Z). We would also like to thank the editor Dr. Carmelo Andújar and the two reviewers (Jan Axtner and anonymous) for providing invaluable feedback, which has significantly improved our manuscript.

References

Akane A, Matsubara K, Nakamura H, Takahashi S, Kimura K (1994) Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. Journal of Forensic Sciences 39(2): 362–372. https://doi.org/10.1520/JFS13607J

Al-Nakeeb K, Petersen T, Sicheritz-Pontén T (2017) Norgal: Extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. BMC Bioinformatics 18: e510. https://doi.org/10.1186/s12859-017-1927-y

Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution 9(1): 134–147. https://doi.org/10.1111/2041-210X.12849

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed 13.10.2021]

Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Virk B, Dalle-Pezze P, Wingett S, Saadeh H, Ahlfors H (2015) Trim Galore. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [accessed 13.10.2021]

Ang A, Roesma DI, Nijman V, Meier R, Srivathsan A, Rizaldi (2020) Faecal DNA to the rescue: Shotgun sequencing of non-invasive samples reveals two subspecies of Southeast Asian primates to be Critically Endangered species. Scientific Reports 10(1): e9396. https://doi.org/10.1038/s41598-020-66007-8

Azam F, Malfatti F (2007) Microbial structuring of marine ecosystems. Nature Reviews. Microbiology 5(10): 782–791. https://doi.org/10.1038/nrmicro1747

Bai Y, Recuenco S, Gilbert AT, Osikowicz LM, Gómez J, Rupprecht C, Kosoy MY (2012) Prevalence and diversity of Bartonella spp. in bats in Peru. The American Journal of Tropical Medicine and Hygiene 87(3): 518–523. https://doi.org/10.4269/ajtmh.2012.12-0097

Bergner LM, Orton RJ, da Silva Filipe A, Shaw AE, Becker DJ, Tello C, Biek R, Streicker DG (2019) Using noninvasive metagenomics to characterize viral communities from wildlife. Molecular Ecology Resources 19(1): 128–143. https://doi.org/10.1111/1755-0998.12946

Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, Biek R, Streicker DG (2020a) Demographic and environmental drivers of metagenomic viral diversity in vampire bats. Molecular Ecology 29(1): 26–39. https://doi.org/10.1111/mec.15250

Bergner LM, Orton RJ, Streicker DG (2020b) Complete Genome Sequence of an Alphacoronavirus from Common Vampire Bats in Peru. Microbiology Resource Announcements 9(34): e00742-20. https://doi.org/10.1128/MRA.00742-20

Bergner LM, Becker DJ, Tello C, Carrera JE, Streicker DG (2021a) Detection of Trypanosoma cruzi in the saliva of diverse neotropical bats. Zoonoses and Public Health 68(3): 271–276. https://doi.org/10.1111/zph.12808

Bergner LM, Mollentze N, Orton RJ, Tello C, Broos A, Biek R, Streicker DG (2021b) Characterizing and Evaluating the Zoonotic Potential of Novel Viruses Discovered in Vampire Bats. Viruses 13(2): e252. https://doi.org/10.3390/v13020252

Bohmann K, Gopalakrishnan S, Nielsen M, Nielsen LDSB, Jones G, Streicker DG, Gilbert MTP (2018) Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. Molecular Ecology Resources 18(5): 1050–1063. https://doi.org/10.1111/1755-0998.12891

Bon C, Berthonaud V, Maksud F, Labadie K, Poulain J, Artiguenave F, Wincker P, Aury J-M, Elalouf J-M (2012) Coprolites as a source of information on the genome and diet of the cave hyena. Proceedings. Biological Sciences 279(1739): 2825–2830. https://doi.org/10.1098/rspb.2012.0358

Calfayan LM, Bonnot G, Villafañe IEG (2018) Case reports of common vampire bats *Desmodus rotundus* (Chiroptera: Phyllostomidae: Desmodontinae) attacking wild exotic mammals in Argentina. Notas Sobre Mamiferos Sudamericanos. https://doi.org/10.31687/saremNMS.19.0.05

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10(1): e421. https://doi.org/10.1186/1471-2105-10-421

Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, Wales N, Sicheritz-Pontén T, Gilbert MTP (2018) Single-tube library preparation for degraded DNA. Methods in Ecology and Evolution 9(2): 410–419. https://doi.org/10.1111/2041-210X.12871

Carter G, Brown B, Razik I, Ripperger S (2021) Penguins, Falcons, and Mountain Lions: The Extraordinary Host Diversity of Vampire Bats. In: Lim BK, Fenton MB, Brigham RM, Mistry S, Kurta A, Gillam EH, Russell A, Ortega J (Eds) 50 Years of Bat Research: Foundations and New Frontiers. Springer International Publishing, Cham, 151–170. https://doi.org/10.1007/978-3-030-54727-1_10

Catenazzi A, Donnelly MA (2008) Sea lion Otaria flavescens as host of the common vampire bat Desmodus rotundus. Marine Ecology Progress Series 360: 285–289. https://doi.org/10.3354/meps07393

Chua PYS, Crampton-Platt A, Lammers Y, Alsos IG, Boessenkool S, Bohmann K (2021) Metagenomics: A viable tool for reconstructing herbivore diet. Molecular Ecology Resources 21(7): 2249–2263. https://doi.org/10.1111/1755-0998.13425

Deagle BE, Eveson JP, Jarman SN (2006) Quantification of damage in DNA recovered from highly degraded samples – a case study in faeces. Frontiers in Zoology 3: e11. https://doi.org/10.1186/1742-9994-3-11

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Molecular Ecology 18(9): 2022–2038. https://doi.org/10.1111/j.1365-294X.2009.04158.x

Delpietro HA, Marchevsky N, Simonetti E (1992) Relative population densities and predation of the common vampire bat (*Desmodus rotundus*) in natural and cattle-raising areas in north-east Argentina. Preventive Veterinary Medicine 14(1–2): 13–20. https://doi.org/10.1016/0167-5877(92)90080-Y

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45: e18. https://doi.org/10.1093/nar/gkw955

Flesher KM, Medici EP (2022) The distribution and conservation status of *Tapirus terrestris* in the South American Atlantic Forest. Neotropical Biology and Conservation 17(1): 1–19. https://doi.org/10.3897/neotropical.17.e71867

Foresman PS (2007) Tapir. Wikimedia Commons, the free media repository. https://commons.wikimedia.org/w/index.php?title=File:Tapir_(PSF)svg&oldid=597619943 [accessed 13.12.2021]

Forouzan E, Maleki MSM, Karkhane AA, Yakhchali B (2017) Evaluation of nine popular de novo assemblers in microbial genome assembly. Journal of Microbiological Methods 143: 32–37. https://doi.org/10.1016/j.mimet.2017.09.008

Galetti M, Pedrosa F, Keuroghlian A, Sazima I (2016) Liquid lunch - vampire bats feed on invasive feral pigs and other ungulates. Frontiers in Ecology and the Environment 14(9): 505–506. https://doi.org/10.1002/fee.1431

Gómez-Rodríguez C, Crampton-Platt A, Timmermans MJT, Baselga A, Vogler AP (2015) Validating the power of mitochondrial metagenomics for community ecology and phylogenetics of complex assemblages. Methods in Ecology and Evolution 6(8): 883–894. https://doi.org/10.1111/2041-210X.12376

Greenhall AM (1970) The use of a precipitin test to determine host preferences of the vampire bats, *Desmodus rotundus* and *Diaemus youngi*. Bijdragen tot de Dierkunde 40(1): 36–39. https://doi.org/10.1163/26660644-04001011

Greenhall AM (1988) Feeding Behavior. In: Greenhall AM, Schmidt RK (Eds) Natural History of Vampire Bats. CRC Press, 7–17. https://books.google.com/books?hl=en&lr=&id=20paDwAAQBAJ&oi=fnd&pg=PA37&dq=greenhall+1988&ots=UQ4m767ZB9&sig=fmcCOX4-Jy44Dn41D5hqixoxZXQ

Greenhall AM, Joermann G, Schmidt U (1983) *Desmodus rotundus*. Mammalian Species. https://doi.org/10.2307/3503895

Hayes MA, Piaggio AJ (2018) Assessing the potential impacts of a changing climate on the distribution of a rabies virus vector. PLoS ONE 13(2): e0192887. https://doi.org/10.1371/journal.pone.0192887

Hoare CA (1965) Vampire Bats as Vectors and Hosts of Equine and Bovine Trypanosomes. Acta Tropica 22(3): 204–216. https://books.google.com/books/about/Vampire_Bats_as_Vectors_and_Hosts_of_Equ.html?hl=&id=YpUBGwAACAAJ

Huson DH, Weber N (2013) Microbial community analysis using MEGAN. Methods in Enzymology 531: 465–485. https://doi.org/10.1016/B978-0-12-407863-5.00021-6

Lee DN, Papeş M, Van den Bussche RA (2012) Present and potential future distribution of common vampire bats in the Americas and the associated risk to cattle. PLoS ONE 7(8): e42466. https://doi.org/10.1371/journal.pone.0042466

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. http://arxiv.org/abs/1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25(16): 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Luna-Jorquera G, Culik BM (1995) Penguins bled by vampires. Journal für Ornithologie 136(4): 471–472. https://doi.org/10.1007/BF01651597

Madden, T (2013) The BLAST sequence analysis tool. The NCBI Handbook. 2nd edn. https://www.ncbi.nlm.nih.gov/sites/books/NBK153387 [accessed 03.11.2021]

Mak SST, Gopalakrishnan S, Carøe C, Geng C, Liu S, Sinding M-HS, Kuderna LFK, Zhang W, Fu S, Vieira FG, Germonpré M, Bocherens H, Fedorov S, Petersen B, Sicheritz-Pontén T, Marques-Bonet T, Zhang G, Jiang H, Gilbert MTP (2017) Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. GigaScience 6(8): 1–13. https://doi.org/10.1093/gigascience/gix049

Mendoza MLZ, Xiong Z, Escalera-Zamudio M, Runge AK, Thézé J, Streicker D, Frank HK, Loza-Rubio E, Liu S, Ryder OA, Samaniego JAC, Katzourakis A, Pacheco G, Taboada B, Löber U, Pybus OG, Li Y, Rojas-Anaya E, Bohmann K, Baez AC, Arias CF, Liu S, Greenwood AD, Bertelsen MF, White EN, Bunce M, Zhang G, Sicheritz-Pontén T, Gilbert MTP (2018) Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. Nature Ecology & Evolution 2(4): 659–668. https://doi.org/10.1038/s41559-018-0476-8

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor protocols 2010: db.prot5448. https://doi.org/10.1101/pdb.prot5448

Nguyen TV, Tilker A, Nguyen A, Hörig L, Axtner J, Schmidt A, Le M, Nguyen AHQ, Rawson BM, Wilting A, Fickel J (2021) Using terrestrial leeches to assess the genetic diversity of an elusive species: The Annamite striped rabbit Nesolagus timminsi. Environmental DNA 3(4): 780–791. https://doi.org/10.1002/edn3.182

Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Detter JC, Pääbo S, Rubin EM (2005) Genomic sequencing of Pleistocene cave bears. Science 309(5734): 597–599. https://doi.org/10.1126/science.1113485

Paula DP, Linard B, Andow DA, Sujii ER, Pires CSS, Vogler AP (2015) Detection and decay rates of prey and prey symbionts in the gut of a predator through metagenomics. Molecular Ecology Resources 15(4): 880–892. https://doi.org/10.1111/1755-0998.12364

Paula DP, Linard B, Crampton-Platt A, Srivathsan A, Timmermans MJTN, Sujii ER, Pires CSS, Souza LM, Andow DA, Vogler AP (2016) Uncovering Trophic Interactions in Arthropod Predators through DNA Shotgun-Sequencing of Gut Contents. PLoS ONE 11(9): e0161841. https://doi.org/10.1371/journal.pone.0161841

Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England) 26(6): 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7(3): 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Reeves LE, Gillett-Kaufman JL, Kawahara AY, Kaufman PE (2018) Barcording blood meals: New vertebrate-specific primer sets for assigning taxonomic identities to host DNA from mosquito blood meals. PLoS Neglected Tropical Diseases 12(8): e0006767. https://doi.org/10.1371/journal.pntd.0006767

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 22(5): 939–946. https://doi.org/10.1101/gr.128124.111

Schneider MC, Romijn PC, Uieda W, Tamayo H, da Silva DF, Belotto A, da Silva JB, Leanes LF (2009) Rabies transmitted by vampire bats to humans: An emerging zoonotic disease in Latin America? Pan. American Journal of Public Health 25: 260–269. https://doi.org/10.1590/S1020-49892009000300010

Sharpton TJ (2014) An introduction to the analysis of shotgun metagenomic data. Frontiers in Plant Science 5: e209. https://doi.org/10.3389/fpls.2014.00209

Srivathsan A, Sha JCM, Vogler AP, Meier R (2015) Comparing the effectiveness of metagenomics and metabarcoding for diet analysis of a leaf-feeding monkey (*Pygathrix nemaeus*). Molecular Ecology Resources 15(2): 250–261. https://doi.org/10.1111/1755-0998.12302

Srivathsan A, Ang A, Vogler AP, Meier R (2016) Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. Frontiers in Zoology 13(1): e17. https://doi.org/10.1186/s12983-016-0150-4

Streicker DG, Allgeier JE (2016) Foraging choices of vampire bats in diverse landscapes: Potential implications for land-use change and disease transmission. Journal of Applied Ecology 53(4): 1280–1288. https://doi.org/10.1111/1365-2664.12690

Tournayre O, Leuchtmann M, Galan M, Trillat M, Piry S, Pinaud D, Filippi-Codaccioni O, Pontier D, Charbonnel N (2021) eDNA metabarcoding reveals a core and secondary diets of the greater horseshoe bat with strong spatio-temporal plasticity. Environmental DNA 3(1): 277–296. https://doi.org/10.1002/edn3.167

Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples–a case study on prey DNA in Antarctic krill stomachs. Frontiers in Zoology 5(12): e12. https://doi.org/10.1186/1742-9994-5-12

Voigt CC, Kelm DH (2006) Host Preference of the Common Vampire Bat (*Desmodus rotundus*; Chiroptera) Assessed by Stable Isotopes. Journal of Mammalogy 87(1): 1–6. https://doi.org/10.1644/05-MAMM-F-276R1.1

Volokhov DV, Becker DJ, Bergner LM, Camus MS, Orton RJ, Chizhikov VE, Altizer SM, Streicker DG (2017) Novel hemotropic mycoplasmas are widespread and genetically diverse in vampire bats. Epidemiology and Infection 145(15): 3154–3167. https://doi.org/10.1017/S095026881700231X

Wilting A, Nguyen TV, Axtner J, Nguyen A, Schmidt A, Le M, Nguyen AHQ, Rawson BM, Tilker A, Fickel J (2021) Creating genetic reference datasets: Indirect sampling of target species using terrestrial leeches as sample "collectors." Environmental DNA. https://doi.org/10.1002/edn3.256

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. Journal of Computational Biology 7(1–2): 203–214. https://doi.org/10.1089/10665270050081478

Zortéa M, Silva DA, Calaça AM (2018) Susceptibility of targets to the vampire bat Desmodus rotundus are proportional to their abundance in Atlantic Forest fragments? Iheringia. Série Zoologia 108: e2018037. https://doi.org/10.1590/1678-4766e2018037

Supplementary material 1

A two-step metagenomics approach for prey identification from the blood meals of common vampire bats (*Desmodus rotundus*)

Author: Physilia Y. S. Chua, Christian Carøe, Alex Crampton-Platt, Claudia S. Reyes-Avila, Gareth Jones, Daniel G. Streicker, Kristine Bohmann

Data type: Supplementary Tables

Explanation note: **Table S1.** Information on where each common vampire bat blood meal sample was collected and corresponding prey taxa analysed using metabarcoding. **Table S2.** Reference seed files used for NOVOPlasty mtDNA contig assembly of common vampire bat blood meal metagenomic samples based on diet items identified from the BLAST-alignment step. **Table S3.** Information on reference seed files used for NOVOPlasty mtDNA contig assembly of diet items identified from common vampire bat blood meals based on the output from the BLAST-alignment step. **Table S4.** Reference seed files used for NOVOPlasty mtDNA contig assembly of common vampire bat blood meal metagenomic samples where no prey was identified from the first BLAST-alignment step. **Table S5.** Reference seed files used for testing the accuracy of NOVOPlasty mtDNA contig assembly by using seed files belonging to species identified from the other samples. **Table S6.** Taxonomic identification of prey from common vampire bat blood meals based on mapping metagenomic reads to COI barcode. The number of reads mapped to the COI barcode is shown for each blood meal sample. **Table S7.** Taxonomic identification of prey from common vampire bat blood meals based on mapping metagenomic reads to COI barcode when using less stringent filtering parameters of 98% sequence identity and 80 bp overlap.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.6.78756.suppl1